Presley Ellsworth

Capitol Technology University

Master of Research (MRes) in Forensic Cyberpsychology

Deepfake Artificial Consciousness: The Ethical, Legal, Psychological, and Forensic Implications of Simulating Machine Consciousness and Synthesizing False Cognitive Emotional Responses

Thesis Chair: Dr. Elena Spiridon, Dr. Pascual Marques

Presley Ellsworth

CTU ID: 135187

December 19, 2024

Table of Contents

Table of Figures

No table of figures entries found.

Table of Tables

Abstract

- A concise summary of the research topic, objectives, and key findings.

- Highlight the integrative approach and the thematic areas explored (ethical, legal, psychological, forensic).

- Mention the relevance of the topic in emerging AI advancements.

The rapid advancement of artificial intelligence and deepfake technologies has enabled the simulation of human-like consciousness and the generation of false cognitive-emotional responses. This integrative literature review explores the multidimensional implications of deepfake artificial consciousness, focusing on its **ethical, legal, psychological, and forensic** ramifications. Ethically, the creation of machines capable of mimicking cognitive and emotional states raises questions about deception, trust, and the moral status of artificially simulated entities. Legally, the lack of robust frameworks to regulate deepfake consciousness and accountability creates challenges in addressing misuse, such as emotional manipulation or misinformation. Psychologically, simulated emotional responses threaten human perception of reality, introducing cognitive dissonance and eroding trust in human-machine interactions. From a forensic perspective, the sophistication of deepfake consciousness complicates the detection of fabricated emotional outputs, posing challenges for digital forensics, cybersecurity, and evidence validation. This review synthesizes current research, identifies gaps in knowledge, and highlights the urgent need for interdisciplinary approaches to address the societal, technological, and regulatory challenges posed by the rise of deepfake artificial consciousness.

Presley Ellsworth

Deepfake Artificial Consciousness: The Ethical, Legal, Psychological, and Forensic Implications of Simulating Machine Consciousness and Synthesizing Deepfaked Cognitive Emotional Responses

1 Introduction

- Background of the study: Briefly introduce **deepfake technology**, machine consciousness, and simulated emotional responses.
- Research problem: The intersection of deepfakes and the perception of artificial consciousness.
- Significance: Why understanding ethical, legal, psychological, and forensic implications is critical in the AI landscape.
- Scope and objectives: Outline the goals of the literature review.
- Research questions guiding the review (e.g., What are the potential dangers of simulating artificial consciousness? How might deepfakes of consciousness challenge societal norms?).

1.1 Background

Artificial intelligence (AI) is advancing rapidly, with deepfake AI and synthetic media impacting entertainment, pornography, and political persuasion (Chesney & Citron, 2019). AI development is trending towards Artificial General Intelligence (AGI) and Artificial Super Intelligence (ASI), which could eventually achieve artificial consciousness (Goertzel, 2014). Current AI, known as Artificial Narrow Intelligence (ANI), already mimics human behavior through sophisticated programming and deep learning (Russell & Norvig, 2021). As AGI progresses, the potential for machine consciousness may emerge (Kurzweil, 2005), and within the next 10-20 years, ASI could surpass human intelligence and display complex emotional and cognitive traits (Bostrom, 2014).

Deepfakes, which create misleading images, videos, and audio, may develop alongside AI (Kietzmann et al., 2020), and when combined with AI's cognitive simulations, it may blur the lines between genuine consciousness and imitation (Tegmark, 2017). Society may struggle to differentiate between deepfake AI consciousness and actual AI awareness (Grover, 2023).

As AI advances, updated ethics, laws, and forensic cyberpsychology frameworks are needed to address the gaps and the growing risk of deepfake artificial consciousness (Floridi & Chiriatti, 2020). The consequences for society, security, and individual identity are profound, particularly as AI begins to simulate consciousness and emotional responses (Whittemore & Knafl, 2005). History shows that ethical and legal frameworks often lag technological advancements (Gunkel, 2018), and deepfake AI is already eroding trust in technology (Kietzmann et al., 2020).

This study will examine the ethical, legal, psychological, and forensic implications of simulating machine consciousness and synthesizing false emotional responses, particularly in the domain of

cybersecurity. Immediate action is needed to update ethical guidelines and regulations before AI has a lasting impact on society, the human psyche and cybersecurity (Gunkel, 2018).

1.2 Significance of Study

The advancement of AI systems that simulate human consciousness present ethical, legal, psychological, and forensic challenges. Current rules and regulations in these fields are inadequate to address the implications of AI-generated emotional and cognitive responses, which blur the line between genuine human behavior and synthesized deepfaked cognitive emotional responses from a machine (Floridi & Chiriatti, 2020; Gunkel, 2018). As AI evolves, there is a growing need to ensure that society can handle deception, misinformation, identity manipulation, and criminal misuse that could result from deepfake artificial consciousness (Floridi & Chiriatti, 2020).

This study is significant because it intends to expose gaps and address updates to ethical guidelines and legal standards, particularly in cybersecurity, which can better address AI systems mimicking human consciousness.

The study will also examine forensic cyberpsychology protocols, which will assist investigators to differentiate between conscious human actions and AI-simulated conscious behaviors. This is critical in cases involving identity fraud and other criminal activities that utilize AI (Gunkel, 2018; Whittemore & Knafl, 2005).

Additionally, the research will explore how deepfake technology and synthetic media can influence public perception of AI as "alive" or sentient. This raises broader psychological and societal concerns. As AI-generated cognitive and emotional responses become more sophisticated, society must have the tools to discern between authentic and artificial interactions to prevent harm and manipulation (Kietzmann, McCarthy, & Silva, 2020).

By investigating these issues, this study will contribute to the development of ethics, laws, regulations, and forensic policy that can help mitigate the potential dangers posed by AI systems simulating consciousness.

Eventually, the goal is to provide a baseline that will help ensure that AI technologies are used ethically and responsibly, while protecting individuals and society from the risks associated with these advanced systems (Floridi & Chiriatti, 2020).

1.3 Purpose of Research (thesis)

The purpose of this integrative literature review research study is to analyze the ethical, legal, psychological, and forensic challenges when deepfake AI consciousness becomes mainstream.

The research will focus on the critical evaluation and gaps in ethics, laws, psychology, and the forensics of deepfake artificial consciousness. These gaps will highlight that when the machines start being conceived as being conscious, they begin to touch on the grey areas of having certain rights and protections (Gunkel, 2018).

Although there are applications and treatments that can benefit by simulating artificial consciousness (e.g., therapy, coaching, Alzheimer's, dementia) (Levy, 2019, Savulescu, 2021) this research endeavors to show that because of the risks and unforeseen negative side effects with simulating artificial consciousness, there needs to be ethics, laws and regulations updated to address this potential threat.

Because there is no consensus on what defines artificial consciousness (Grover, 2023), this research will also outline what constitutes human consciousness and how that outline serves as a guide towards simulating consciousness in an AI system (Chalmers, 1996).

Potential hypotheses:

- Simulating machine consciousness with deepfake technology raises ethical issues, particularly the intentional deception and exploitation of human users.
- Currently there are insufficient laws and regulations to address the impact that deepfake AI could have on property, identity theft, and psychological wellbeing.
- AI systems that simulate consciousness and synthesize emotional responses and cognitive ability yield measurable psychological effects on users.
- Deepfake technologies that simulate consciousness in AI systems create unique challenges for forensic investigators' ability to distinguish between human and AI-generated behaviors in digital evidence.
- The use of deepfake technology on advanced AI systems like AGI and ASI undermines the trust and reputation of an AI system, as users may not be able to discern the difference between human reaction and AI-generated reactions.

1.4 General Problem Statement

AI technologies are advancing at rapid rates. AI is already responding to prompts with personable tones and friendliness. When deepfakes and chatbots are programmed to respond in an artificially conscious way, there will be a threat to ethical, legal psychological and forensic frameworks (Gunkel, 2018; Savulescu, 2021). Current frameworks are ill equipped to deal with the challenges posed by AI systems that mimic cognitive and emotional behaviors. There is a gap that is emerging in how these systems will impact human behavior, legal doctrine, and security mechanisms (Floridi & Chiriatti, 2020).

1.5 Specific Problem Statement

AI technologies are advancing at rapid rates. AI is already responding to prompts with personable tones and friendliness. When deepfakes and chatbots are programmed to respond in an artificially conscious way, there will be a threat to ethical, legal psychological and forensic frameworks (Gunkel, 2018; Savulescu, 2021). Current frameworks are ill equipped to deal with the challenges posed by AI systems that mimic cognitive and emotional behaviors. There is a gap that is emerging in how these systems will impact human behavior, legal doctrine, and security mechanisms (Floridi & Chiriatti, 2020).

1.6 Research Questions:

1. What ethical policies, standards and regulations are currently in place to address AI systems simulating human consciousness, and
    a. How can these policies, standards and regulations be updated to address emerging challenges?
2. What are the legal implications when AI-generated emotional and cognitive responses involve identity manipulation and criminal behavior?
3. What forensic techniques can be developed, adapted, or manipulated to help differentiate between human behavior and AI-generated simulations of consciousness?
4. What are the vulnerabilities and risks created by an AI system capable of imitating human consciousness?
    a. How can these vulnerabilities and risks be mitigated?
5. Which existing laws and regulations can be modified to address the deception and misuse of AI systems that mimic consciousness?

2 Conceptual Framework

- Define key concepts and terms:

    o **Deepfake technology**: Overview of its mechanisms, including AI models like GANs.

    o **Artificial consciousness**: Defining the simulation of cognition, consciousness, and emotional states.

    o **False cognitive-emotional responses**: Deepfakes simulating emotional states and perceptions.

- **Theoretical approaches: Introduce relevant frameworks, e.g., technological determinism, ethical AI theories, and cyberpsychological models.**

This research will blend the disciplines of forensic cyberpsychology, AI, ethics, legislation and cognitive science together to understand how AI systems can manipulate human-machine interactions. This conceptual framework can also be used to understand how forensic cyberpsychologists can detect and assess AI systems that are simulating artificial consciousness.

The stage of the framework for an integrative literature review for this thesis is as follows:

1. Problem Formulation

- Formulate clear research questions or objectives (Torraco, 2005).
- Establish inclusion and exclusion criteria for selecting literature (Whittemore & Knafl, 2005).
- Conduct a comprehensive search of literature using multiple databases (e.g., ACM Digital Library, EBSCOHost Databases and Services, Google Scholar, IEEE Xplore, ProQuest One Academic) (Booth, Sutton, & Papaioannou, 2016).

2. Literature Search

- Use relevant keywords and search terms (Torraco, 2016).
- Set criteria for inclusion and exclusion, ensuring the comprehensiveness of the search (Whittemore & Knafl, 2005).

3. Data Evaluation

- Categorize studies based on their methodologies, theoretical contributions, and practical relevance (Torraco, 2016).
- Evaluate the rigor of research methodologies used in the selected literature (Whittemore & Knafl, 2005).
- Critically appraise the methodologies and findings of each study (Torraco, 2016).

4. Data Analysis and Synthesis

- Synthesize results by combining insights from multiple sources to form new conceptual frameworks (Torraco, 2005).
- Develop a narrative or framework to explain how different studies contribute to the understanding of the topic (Whittemore & Knafl, 2005).
- Identify patterns, themes, and gaps in the literature (Russell, 2005).

5. Presentation of Findings

- Structure the review around key themes or concepts that emerged during synthesis (Torraco, 2005).
- Provide a cohesive discussion of the implications of the findings for theory, practice, and future research (Russell, 2005).
- Clearly present the critical analysis of the literature, addressing research gaps and limitations (Torraco, 2016).

6. Conclusion and Recommendations

- Summarize the current state of knowledge on the topic (Torraco, 2016).
- Identify limitations in the reviewed literature (Whittemore & Knafl, 2005).
- Suggest future research directions based on gaps in the existing literature (Russell, 2005).

2.1 Preliminary Proposed Research Methods

This study will employ an integrative literature review methodology as the primary research method to explore the ethical, legal, psychological, and forensic implications of simulating machine consciousness and synthesizing false cognitive emotional responses. Integrative reviews are well-suited for synthesizing diverse types of research (e.g., theoretical, empirical) to generate new insights into emerging fields like artificial intelligence (Torraco, 2005; Whittemore & Knafl, 2005).

2.2 Research Design

The integrative literature review will adhere to established guidelines for conducting and synthesizing research. This approach will include the following stages: problem formulation, literature search, data evaluation, data analysis and synthesis, and presentation of findings (Torraco, 2005; Whittemore & Knafl, 2005). The review will utilize peer-reviewed articles, books, conference proceedings, and scholarly reports that address the intersection of deepfake technology and AI consciousness. The ethics, legal frameworks, and forensic implications will be assessed and gaps in these areas will be identified.

2.3 Data Sources and Search Strategy

A comprehensive literature search will be conducted across academic databases (e.g., Google Scholar, IEEE Xplore, ACM Digital Library, Academia.edu, Researchgate.com, ProQuest) to gather relevant studies. Keywords and phrases such as "AI consciousness ethics," "deepfake AI consciousness," "simulated machine consciousness," "legal implications of AI consciousness," and "forensic AI consciousness" will guide the search (Gunkel, 2018; Levy, 2019; Savulescu, 2021). To ensure the relevance and currency of the review, only sources from the last ten years

will be included, unless older works, such as those by Chalmers (1996), contribute significantly to the study.

2.4 Inclusion and Exclusion Criteria

The inclusion criteria for this review will focus on:

- **Relevance**: Studies that specifically address the ethical, legal, psychological, and forensic implications of AI systems simulating human consciousness.

- **Type of Study**: The review will include both empirical and theoretical research, as integrative reviews benefit from a combination of data sources (Whittemore & Knafl, 2005).

- **Quality**: Only peer-reviewed and academically credible sources will be considered.

Exclusion criteria will include non-academic articles, opinion pieces without empirical or theoretical backing, and studies that do not directly address the topic of AI consciousness.

2.5 Data Analysis

The data will be evaluated using thematic analysis to identify key themes and gaps in the existing literature (Braun & Clarke, 2006). Thematic analysis is appropriate for synthesizing findings from a diverse set of sources and is widely used in integrative reviews to organize and interpret data (Torraco, 2005). Key themes, such as the ethical challenges of deepfake AI, legal issues surrounding AI-generated cognitive responses, and forensic difficulties in distinguishing human from AI behaviors, will be identified and discussed.

2.6 Ethical Considerations

While this study is based on a literature review and does not involve primary data collection from human subjects, ethical considerations include ensuring that the review is conducted rigorously and transparently, with careful attribution of sources. The study will adhere to guidelines for ethical research and review practices, including maintaining objectivity, avoiding plagiarism, and respecting the intellectual property of the authors reviewed (Booth, Sutton, & Papaioannou, 2016).

3 Literature Review

3.1 Deepfakes and Artificial Consciousness

- Overview of deepfake evolution and its expansion into cognitive/emotional simulation.
- The distinction between true machine consciousness and simulated consciousness.
- Current research on how deepfake AI can imitate emotional and cognitive behavior.

3.2 Ethical Implications

- Ethical dilemmas of simulating machine consciousness:
- Deception and false emotional connections.
- Implications of creating machines that mimic cognitive and emotional states.
- Philosophical concerns:
- Can machines achieve moral or ethical consideration if consciousness is faked?
- Risks of misuse: Manipulation, misinformation, and human emotional exploitation.

3.3 Legal Implications

- Current legal frameworks addressing deepfakes (e.g., privacy laws, intellectual property).
- Gaps in legislation regarding AI and simulated consciousness.
- Accountability and liability:
- Who is responsible for harm caused by deepfake consciousness?
- Case studies of deepfake misuse in legal systems or public trust.

3.4 Psychological Implications

- Impact on human perception of reality:
- Cognitive dissonance and trust erosion.
- Psychological responses to interacting with AI mimicking emotions.
- Emotional manipulation and exploitation:
- How fake cognitive responses influence behavior, relationships, and decision-making.
- The uncanny valley effect: Psychological discomfort with highly realistic yet "non-conscious" AI.

3.5 Forensic Implications

- Challenges in detecting deepfakes simulating consciousness and emotions.
- Forensic tools and technologies for identifying AI-generated emotional simulations.
- Implications for digital forensics:

- How deepfake consciousness affects evidence authenticity.
- Legal proceedings and investigations involving deceptive AI.
- Future needs in forensic AI detection and ethical technology standards.

4 Discussion

- Synthesize the literature across the four thematic areas: ethical, legal, psychological, and forensic implications.
- Identify common threads and key gaps in the literature.
- Highlight areas where interdisciplinary approaches are needed (e.g., technology, law, psychology, ethics).
- Explore the broader societal impact:
    - How could deepfake artificial consciousness disrupt trust, identity, and communication?
    - Potential benefits of simulated consciousness for therapy, entertainment, or communication (counterbalancing risks).

5 Recommendations

- **For Ethics**: Propose ethical guidelines for AI developers and researchers simulating emotional states.
- **For Law**: Suggest updates to legal frameworks to address deepfake consciousness and accountability gaps.
- **For Psychology**: Advocate for public awareness programs about deepfake emotional manipulation and psychological resilience.
- **For Forensics**: Highlight the need for advanced detection tools and training for digital forensic experts.

6 Conclusion

- Summarize key findings and contributions of the review.
- Reiterate the critical nature of addressing deepfake artificial consciousness.
- Discuss future directions for research, policy, and interdisciplinary collaboration.

7 References

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies.* Oxford University Press.

Booth, A., Sutton, A., & Papaioannou, D. (2016). *Systematic approaches to a successful literature review* (2nd ed.). SAGE Publications.

https://www.researchgate.net/profile/Andrew-Booth-2/publication/235930866_Systematic_Approaches_to_a_Successful_Literature_Review/links/5da06c7f45851553ff8705fa/Systematic-Approaches-to-a-Successful-Literature-Review.pdf

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa; https://www.researchgate.net/publication/235356393_Using_thematic_analysis_in_psychology

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford University Press.

Chesney, R., & Citron, D. K. (2019). Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. *Foreign Affairs, 98*(1), 147-155. https://perma.cc/TW6Z-Q97D

Coeckelbergh, M. (2020). AI ethics. *The MIT Press Essential Knowledge Series.* MIT Press. https://www.researchgate.net/publication/339103412_AI_Ethics#fullTextFileContent

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30(4), 681-694. https://doi.org/10.1007/s11023-020-09548-1

Goertzel, B. (2014). Artificial general intelligence: Concept, state of the art, and future prospects. *Journal of Artificial General Intelligence, 5*(1), 1-46. https://www.researchgate.net/publication/271390398_Artificial_General_Intelligence_Concept_State_of_the_Art_and_Future_Prospects#fullTextFileContent

Grover, M. (2023). Artificial Consciousness And The Universal Field Of Consciousness, *International Journal of Research and Analytical Reviews,* 10(1), E-ISSN 2348-1269, P-ISSN 2349-5138, https://www.academia.edu/97663146/Artificial_Consciousness_And_The_Universal_Field_Of_Consciousness

Gunkel, D. J. (2018). Robot rights. The MIT Press. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4077131

Kietzmann, J., McCarthy, I., & Silva, H. (2020). Deepfakes: Trick or treat? *Business Horizons, 63*(2), 135-146. https://www.researchgate.net/profile/Jan-Kietzmann/publication/338144721_Deepfakes_Trick_or_treat/links/5e6aaf64a6fdccf321d91d43/Deepfakes-Trick-or-treat.pdf

Levy, N. (2019). Artificial agents and the expanding moral circle. *Ethics and Information Technology*, *21*(3), 163–175. https://doi.org/10.1007/s10676-019-09505-4; https://www.researchgate.net/publication/257333998_Artificial_agents_and_the_expanding_ethical_circle

Russell, C. L. (2005). An overview of the integrative research review. Progress in Transplantation, 15(1), 8-13. http://dx.doi.org/10.7182/prtr.15.1.0n13660r26g725kj; https://www.researchgate.net/publication/7898657_An_overview_of_the_integrative_research_review

Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson. https://doi.org/10.1109/MSP.2017.2765202

Savulescu, J. (2021). Simulating minds: AI in therapy and beyond. *Journal of Medical Ethics*, *47*(9), 597-601. https://doi.org/10.1136/medethics-2020-107088; https://www.researchgate.net/publication/378711445_AI_and_the_need_for_justification_to_the_patient

Tegmark, M. (2017). *Life 3.0: Being human in the age of artificial intelligence.* Knopf. https://www.cag.edu.tr/uploads/site/lecturer-files/max-tegmark-life-30-being-human-in-the-age-of-artificial-intelligence-alfred-a-knopf-2017-aTvn.pdf

Torraco, R. J. (2016). Writing integrative reviews of the literature: Methods and purposes. *International Journal of Adult Vocational Education and Technology, 7*(3), 62-70. https://doi.org/10.4018/IJAVET.2016070106; https://sageprofessor.com/wp-content/uploads/2017/10/writing-integrative-reviews-of-the-literature-methods-and-purposes.pdf

Whittemore, R., & Knafl, K. (2005). The integrative review: Updated methodology. *Journal of Advanced Nursing, 52*(5), 546-553. http://dx.doi.org/10.1111/j.1365-2648.2005.03621.x; https://www.researchgate.net/publication/7498980_The_integrative_review_Update_methodology

8 Glossary of Terms

Table 1 Glossary of Terms

Term	Definition

9 Acronyms

Table 2 Acronyms

Abbreviation	Definition

10 Key Findings Across Studies

Table 3 Key Findings Across Studies

Term	Definition